

The Exact Schema Theorem

Alden H. Wright

Computer Science Department

University of Montana

Missoula, MT 59812

alden.wright@umontana.edu

<http://web-dev.cs.umt.edu/~wright/wright.htm>

January 21, 2013

Comment added on May 18, 2011: This paper was written in 1999 with the last revision done on January 28, 2000. A colleague dissuaded me from submitting it for publication, but it does contain useful results. It has been published on my website (formerly <http://www.cs.umt.edu/u/wright>) since that 1999. It has been referenced by other publications.

On May 18, 2011, I reprocessed the LaTeX source using LaTeX, dvips, and ps2pdf. (There were some fractions that did not display correctly.) Except for the addition of this comment, updating my e-mail and website URL, and the reprocessing, there have been no changes since the version of January 28, 2000.

Abstract

A schema is a naturally defined subset of the space of fixed-length binary strings. The Holland Schema Theorem [Hol75] gives a lower bound on the expected fraction of a population in a schema after one generation of a simple genetic algorithm. This paper gives formulas for the exact expected fraction of a population in a schema after one generation of the simple genetic algorithm.

Holland's schema theorem has three parts, one for selection, one for crossover, and one for mutation. The selection part is exact, whereas the crossover and mutation parts are approximations. This paper shows how the crossover and mutation parts can be made exact. Holland's schema theorem follows naturally as a corollary.

There is a close relationship between schemata and the representation of the population in the Walsh basis. This relationship is used in the derivation of the results, and can also make computation of the schema averages more efficient.

This paper gives a version of the Vose infinite population model where crossover and mutation are separated into two functions rather than a single "mixing" function.

1 Introduction

Holland's schema theorem [Hol75] has been widely used for the theoretical analysis of genetic algorithms. However, it has two limitations. First, it only gives information for a single generation. Second, it is an approximation, giving only lower bounds on the schema frequencies. This paper removes the second limitation.

Michael Vose and coworkers have introduced exact models of the simple genetic algorithm. The Vose infinite population model exactly describes the expected behavior from one generation to the next. The Markov chain model is an exact model of the finite population behavior of the simple GA.

Stephens et. al. [SW97] describe these models as “fine-grained”. They can be used to qualitatively describe certain aspects of the behavior of the simple GA. For example, the fixed points of the infinite population model can be used to describe phenomena such as punctuated equilibria. (See [VL91] and [Vos99a] for example.) However, due to the large size of the models, it is generally impossible to apply these models quantitatively to practical-sized problems.

Thus, as is pointed out in [SW97] and [SWA97], a more coarse-grained version of these models is needed. Models are needed that describe the behavior of a subset of the variables included in the exact models. For example, a higher-level organism may have in the order of magnitude of 100,000 genes. However, population geneticists generally do not try to model all of these; instead they may use 1-locus and 2-locus models. Modeling using schemata is the equivalent technique for string-representation genetic algorithms; they model the behavior of the GA at a subset of the string positions.

In earlier work, Bridges and Goldberg [BG87] derived an exact expression for expected number of copies of a string under one generation of selection and one-point crossover, and they claim that their formulas can be extended to find the expected number of elements in a schema under the same conditions. Their formulas are complex and not particularly illuminating.

As mentioned before, Stephens and coworkers ([SW97] and [SWA97]) have results similar to ours for one-point crossover. Our results are more general than these results in that they for general crossover, and they include mutation.

[SW97] includes references to other related papers. Of particular note is [Alt95] which relates an exact version of the schema theorem to Price's theorem in population genetics.

Chapter 19 of [Vos99b] (which the author had not seen when he wrote this paper) also contains a version of the exact schema theorem as theorem 19.2 for mixing, where mixing includes crossover and mutation. Theorem 19.2 assumes that mutation is independent, which is similar to the assumptions on mutation in this paper.

2 Notation

Let Ω be the space of length ℓ binary strings, and let $n = 2^\ell$. For $u, v \in \Omega$, let $u \otimes v$ denote the bitwise-and of u and v , and let $u \oplus v$ denote the bitwise-xor of u and v . Let \bar{u} denote the ones-complement of u , and $\#u$ denote the number of ones in the binary representation of u .

Integers in the interval $[0, n) = [0, 2^\ell)$ are identified with the elements of Ω through their binary representation. This correspondence allows Ω to be regarded as the product group

$$\Omega = Z_2 \times \dots \times Z_2$$

where the group operation is \oplus . The elements of Ω corresponding to the integers 2^i , $i = 0, \dots, \ell - 1$ form a natural basis for Ω .

We will also use column vectors of length ℓ to represent elements of Ω . Let $\mathbf{1}$ denote the vector of ones (or the integer $2^\ell - 1$). Thus, $u^T v = \#(u \otimes v)$, and $\bar{u} = \mathbf{1} \oplus u$.

For any $u \in \Omega$, let Ω_u denote the subgroup of Ω generated by $\langle 2^i : u \otimes 2^i = 2^i \rangle$. In other words, $v \in \Omega_u$ if and only if $v \otimes u = v$. For example, if $\ell = 6$, then $\Omega_9 = \{0, 1, 8, 9\} = \{000000, 000001, 001000, 001001\}$.

A schema is a subset of Ω where some string positions are specified (fixed) and some are unspecified (variable). Schemata are traditionally denoted by pattern strings, where a special symbol is used to denote a unspecified bit. We use the $*$ symbol for this purpose (Holland used the $\#$ symbol). Thus, the schema denoted by the pattern string $10*01*$ is the set of strings $\{100010, 100011, 101010, 101011, \dots\}$.

Alternatively, we can define a schema to be the set $\Omega_u \oplus v$, where $u, v \in \Omega$, and where $u \otimes v = 0$. In this notation, u is a mask for the variable positions, and v specifies the fixed positions. For example, the schema $\Omega_{001001} \oplus 100010$ would be the schema $10*01*$ described above.

This definition makes it clear that a schema $\Omega_u \oplus v$ with $v = 0$ is a subgroup of Ω , and a schema $\Omega_u \oplus v$ is a coset of this subgroup.

Following standard practice, we will define the *order* of a schema as the number of fixed positions. In other words, the order of the schema $\Omega_{\bar{u}} \oplus v$ is $\#u$ (since u is a mask for the fixed positions).

A population for a genetic algorithm over length ℓ binary strings is usually interpreted as a multiset (set with repetitions) of elements of Ω . A population can also be interpreted as a 2^ℓ dimensional incidence vector over the index set Ω : if X is a population vector, then X_i is the number of occurrences of $i \in \Omega$ in the population. A population vector can be normalized by dividing by the population size. For a normalized population vector x , $\sum_i x_i = 1$. Let

$$\Lambda = \{x \in R^n : \sum_i x_i = 1 \text{ and } x_i \geq 0 \text{ for all } i \in \Omega\}.$$

Thus a normalized population vector is an element of Λ . Geometrically, Λ is the $n - 1$ dimensional unit simplex in R^n . Note that elements of Λ can be interpreted as probability distributions over Ω .

If $expr$ is a Boolean expression, then

$$[expr] = \begin{cases} 1 & \text{if } expr \text{ is true} \\ 0 & \text{if } expr \text{ is false} \end{cases}$$

3 The fraction of a population in a schema

Let X be a population (not necessarily normalized). We will be interested in the fraction $X_k^{(u)}$ of the elements of X that are elements of the schema $\Omega_{\bar{u}} \oplus k$:

$$X_k^{(u)} = \frac{\sum_{i \in \Omega_{\bar{u}}} X_{i \oplus k}}{\sum_{i \in \Omega} X_i} \quad \text{for } k \in \Omega_u.$$

Note that here u is a mask for the fixed positions of the schema.

If we divide the numerator and denominator of this fraction by the population size r , and if we let $x = X/r$, then we get

$$x_k^{(u)} = \frac{\sum_{i \in \Omega_{\bar{u}}} x_{i \oplus k}}{\sum_{i \in \Omega} x_i} = \sum_{i \in \Omega_{\bar{u}}} x_{i \oplus k}$$

In other words, for a normalized population x , we use the notation $x_k^{(u)}$ to denote the *schema average* for the schema $\Omega_{\bar{u}} \oplus k$. Note that $x_0^{(0)} = 1$ since $\sum_{i \in \Omega} x_i = 1$.

Let $x^{(u)}$ denote the vector of schema averages, where the vector is indexed over Ω_u . Note that $\sum_{v \in \Omega_u} x_v^{(u)} = 1$.

For a fixed u , the family of schemata $\{\Omega_{\bar{u}} \oplus v : v \in \Omega_u\}$ is called a *competing family* of schemata.

4 The Simple Genetic Algorithm

The material in this section is mostly taken from [Vos99b], [Vos96], and [VW98a].

The simple genetic algorithm can be described through a heuristic function $\mathcal{G} : \Lambda \rightarrow \Lambda$. As we will show later, \mathcal{G} contains all of the details of selection, crossover, and mutation. The simple genetic algorithm is given by:

- 1 Choose a random population of size r from Ω .
- 2 Express the population as an incidence vector X indexed over Ω .
- 3 Let $y = \mathcal{G}(X/r)$. (Note that X/r and y are probability distributions over Ω .)
- 4 **for** k from 1 to r **do**
- 5 Select individual $i \in \Omega$ according to the probability distribution y .
- 6 Add i to the next generation population Z .

```

7  endfor
8  Let  $X = Z$ .
9  Go to step 3.

```

It is shown in [Vos99b] that if X is a population, then $y = \mathcal{G}(X/r)$ is the expected population after one generation of the simple genetic algorithm. Thus, the schema theorem is a statement about the schema averages of the population y .

The heuristic function \mathcal{G} can be written as the composition of three heuristic functions \mathcal{F} , \mathcal{C} , and \mathcal{U} which describe selection, crossover, and mutation respectively. In other words, $\mathcal{G}(x) = \mathcal{U}(\mathcal{C}(\mathcal{F}(x))) = \mathcal{U} \circ \mathcal{C} \circ \mathcal{F}(x)$. Later sections describe each of the three heuristic functions in more detail.

5 Selection

The selection heuristic \mathcal{F} for proportional selection is given by:

$$\mathcal{F}_k(x) = \frac{f_k x_k}{\sum_{j \in \Omega} f_j x_j}$$

where f_k denotes the fitness of $k \in \Omega$.

Let F denote the diagonal matrix over $\Omega \times \Omega$ whose diagonal entries are given by $F_{j,j} = f_j$. Then the selection heuristic can be expressed in terms of matrices by

$$\mathcal{F}(x) = \frac{Fx}{\mathbf{1}^T Fx}$$

If X is a finite population represented as an incidence vector over Ω , and if $x = X/r$, then x_k is nonzero only for those k that are in the population X considered as a multiset. Thus, the computation of $\mathcal{F}(x)$ is feasible in practice even for long string lengths. Further, the computation of the schema averages after selection can be done directly from the definition.

Theorem 5.1 (*Exact schema theorem for proportional selection.*) *Let $x \in \Lambda$ be a population, and let $s = \mathcal{F}(x)$. Then*

$$s_k^{(u)} = \frac{\sum_{j \in \Omega_u} f_{j \oplus k} x_{j \oplus k}}{\sum_{k \in \Omega} f_k x_k}$$

We give the following algorithm for computing the schema average vector $s^{(u)}$ from a finite population X . Let $I(u) = \{i : 0 \leq i < \ell \text{ and } u_i = 1\}$, where u_i denotes bit i of u . Let $P^{(u)}$ be the function which projects Ω into Ω_u : for $j \in \Omega$, let $P_i^{(u)}(j) = j_i$ for $i \in I(u)$.

```

for each  $k \in \Omega_u$  do
   $s_k^{(u)} \leftarrow 0$ 
endfor
for each  $j \in X$  do ▷ see note below
   $k \leftarrow P^{(u)}(j)$ 
   $s_k^{(u)} \leftarrow s_k^{(u)} + f_j$ 
endfor
 $\bar{f} \leftarrow 0$ 
for each  $k \in \Omega_u$  do
   $\bar{f} \leftarrow \bar{f} + s_k^{(u)}$ 
endfor
for each  $k \in \Omega_u$  do
   $s_k^{(u)} \leftarrow s_k^{(u)} / \bar{f}$ 
endfor
return  $s_k^{(u)}$ 

```

In this algorithm, the population X is interpreted as a multiset. Thus, it is assumed that “**for each** $j \in X$ **do**” means that the loop following is done once for each of the possibly multiple occurrences of j in X . In an implementation, it would be useful to identify the elements of Ω_u with the integers in the interval $[0, 2^{\#u})$, and to interpret $s^{(u)}$ as a vector indexed over these integers.

Clearly, the complexity of this algorithm is $\Theta(2^{\#u} + rK)$, where K denotes the complexity of one fitness evaluation.

We now give an example which we will continue through the remaining sections.

Let $\ell = 5$, $u = 10 = 01010_2$, $r = 5$, $X = \{6, 7, 10, 13, 21\} = \{00110, 00111, 01010, 01101, 10101\}$. The schema sum vector is $x^{(10)} = \langle \frac{1}{5}, \frac{2}{5}, \frac{1}{5}, \frac{1}{5} \rangle$. Let $f_6 = 5$, $f_7 = 3$, $f_{10} = 4$, $f_{13} = 1$, $f_{21} = 7$. This gives $\bar{f} = 20$. The schema sum vector after selection is $s^{(10)} = \frac{1}{20} \langle 7, 8, 1, 4 \rangle$.

6 Holland’s Schema Theorem

We can now state Holland’s Schema theorem [Hol75].

As in [VW98a], for $u \in \Omega$, define

$$\begin{aligned}
 \text{hi}(u) &= \begin{cases} 0 & \text{if } u = 0 \\ \max\{i : 2^i \otimes u > 0\} & \text{otherwise} \end{cases} \\
 \text{lo}(u) &= \begin{cases} \ell - 1 & \text{if } u = 0 \\ \min\{i : 2^i \otimes u > 0\} & \text{otherwise} \end{cases}
 \end{aligned}$$

Intuitively, the function $\text{hi}(u)$ returns the index high-order bit of u , and $\text{lo}(u)$ returns the index of the low-order bit. Let $\mathcal{L}(u) = \text{hi}(u) - \text{lo}(u)$. $\mathcal{L}(u)$ is often called the defining length of u .

Theorem 6.1 (*Holland's approximate schema theorem.*) *Let $x \in \Lambda$ be a normalized population, and let $y = \mathcal{G}(x)$, where \mathcal{G} includes proportional selection, one-point crossover with crossover rate c , and bitwise mutation with mutation rate p . Then,*

$$y_v^{(u)} \geq \frac{\sum_{j \in \Omega_{\bar{u}}} f_{j \oplus k} x_{j \oplus k}}{\sum_{j \in \Omega} f_j x_j} \left(1 - c \frac{\mathcal{L}(u)}{\ell - 1}\right) (1 - p)^{\#u}$$

7 The Walsh Basis

The Walsh matrix W has dimension 2^ℓ by 2^ℓ , and has elements defined by

$$W_{i,j} = 2^{-\ell/2} (-1)^{i^T j} = \frac{1}{\sqrt{n}} (-1)^{i^T j}$$

Note that W is symmetric and orthogonal ($WW = I$). The columns of W define a basis for R^n called the Walsh basis.

As an example, for $\ell = 2$,

$$W = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

If x is a vector over Ω , then $\hat{x} = Wx$ can be interpreted as x written in the Walsh basis, and if M is a matrix over $\Omega \times \Omega$, then $\widehat{M} = WMW$ can be interpreted as M written in the Walsh basis.

We are also interested in vectors and matrices indexed over Ω_u . If \hat{x} is a vector over Ω written in the Walsh basis, let $\hat{x}_k^{(u)} = 2^{\#\bar{u}/2} \hat{x}_k$. Theorem 7.1 will show that $\hat{x}^{(u)}$ is the Walsh transform of $x^{(u)}$.

We can define a Walsh matrix $W^{(u)}$ indexed over $\Omega_u \times \Omega_u$. For $i, j \in \Omega_u$, define

$$W_{i,j}^{(u)} = 2^{-\#u/2} (-1)^{i^T j}$$

The following theorem shows how the schema sum vector is related to the Walsh coefficients of the population.

Theorem 7.1 *For any $u \in \Omega$,*

$$x^{(u)} = W^{(u)} \hat{x}^{(u)}$$

Proof.

$$\begin{aligned}
(W^{(u)}\hat{x}^{(u)})_k &= 2^{-\#u/2} \sum_{j \in \Omega_u} (-1)^{j^T k} \hat{x}_j^{(u)} \\
&= 2^{\#\bar{u}/2 - \#u/2} \sum_{j \in \Omega_u} (-1)^{j^T k} \hat{x}_j \\
&= 2^{\#\bar{u}/2 - \#u/2} 2^{-\ell/2} \sum_{j \in \Omega_u} (-1)^{j^T k} \sum_{v \in \Omega} (-1)^{j^T v} x_v \\
&= 2^{-\#u} \sum_{v \in \Omega} x_v \sum_{j \in \Omega_u} (-1)^{j^T (v \oplus k)} \\
&= 2^{-\#u} \sum_{w \in \Omega} x_{w \oplus k} \sum_{j \in \Omega_u} (-1)^{j^T w} \tag{1} \\
&= 2^{-\#u} \sum_{w \in \Omega_{\bar{u}}} x_{w \oplus k} \sum_{j \in \Omega_u} (-1)^{j^T w} \tag{2} \\
&= 2^{-\#u} \sum_{w \in \Omega_{\bar{u}}} x_{w \oplus k} 2^{\#u} \\
&= x_k^{(u)}
\end{aligned}$$

To see how equation 2 follows from 1, note that $\sum_{j \in \Omega_u} (-1)^{j^T v} = 0$ if $v \in \Omega_u$ and $v \neq 0$. ■

Theorem 7.1 shows that the schema averages of a family of competing schemata determine the Walsh coefficients of the population in a coordinate subspace in the Walsh basis. To be more specific, consider u as fixed. Then $x^{(u)}$ denotes the schema averages of the family of competing schemata $\Omega_{\bar{u}} + k$, where k varies over Ω_u . Theorem 7.1 shows that these schema averages determine $\hat{x}^{(u)}$, which is a rescaling of the projection of \hat{x} into the coordinate subspace generated by the elements of Ω_u .

To continue the example started in section 5, if $s^{(10)} = \frac{1}{20} < 7, 8, 1, 4 >$, then $\hat{s}^{(10)} = W^{(10)}s^{(10)} = \frac{1}{20} < 10, -2, 5, 1 >$.

8 Crossover

If parent strings $i, j \in \Omega$ are crossed using a crossover mask $m \in \Omega$, the children are $(i \otimes m) \oplus (j \otimes \bar{m})$ and $(i \otimes \bar{m}) \oplus (j \otimes m)$. In the simple genetic algorithm, one child is chosen randomly from the pair of children.

For each binary string $m \in \Omega$, let χ_m be the probability of using m as a crossover mask.

The *crossover matrix* is given by

$$C_{i,j} = \sum_{m \in \Omega} \frac{\chi_m + \chi_{\bar{m}}}{2} [i \otimes m \oplus j \otimes \bar{m} = 0]$$

$C_{i,j}$ is the probability of obtaining 0 as the result of the crossover of i and j .

Let σ_k be the permutation matrix with i, j th entry given by $[j \oplus i = k]$. Then $(\sigma_k x)_i = x_{i \oplus k}$. Define the *crossover heuristic* $\mathcal{C} : \Lambda \rightarrow \Lambda$ by

$$\mathcal{C}_k(x) = (\sigma_k x)^T C \sigma_k x \quad \text{for } k \in \Omega$$

Corollary 3.3 of [VW98a] gives that the Walsh transform of the crossover matrix \hat{C} is equal to C .

Vose and Wright [VW98a] show that the k th component of $\mathcal{C}(x)$ with respect to the Walsh basis is

$$\sqrt{n} \sum_{i \in \Omega_k} \hat{x}_i \hat{x}_{i \oplus k} \hat{C}_{i, i \oplus k}$$

where $n = 2^\ell$.

Theorem 8.1 (*The crossover heuristic in the Walsh basis.*) Let $x \in \Lambda$ and let \hat{y} denote $\mathcal{C}(x)$ expressed in the Walsh basis. Then

$$\hat{y}_k = \sqrt{n} \sum_m \frac{\chi_m + \chi_{\bar{m}}}{2} \hat{x}_{k \otimes m} \hat{x}_{k \otimes \bar{m}}.$$

Proof.

$$\begin{aligned} \hat{y}_k &= \sqrt{n} \sum_{i \in \Omega_k} \hat{x}_i \hat{x}_{i \oplus k} C_{i, i \oplus k} \\ &= \sqrt{n} \sum_{i \in \Omega_k} \hat{x}_i \hat{x}_{i \oplus k} \sum_m \frac{\chi_m + \chi_{\bar{m}}}{2} [(i \otimes m = 0) \wedge ((i \oplus k) \otimes \bar{m} = 0)] \\ &= \sqrt{n} \sum_m \frac{\chi_m + \chi_{\bar{m}}}{2} \sum_{i \in \Omega_k} \hat{x}_i \hat{x}_{i \oplus k} [(i \otimes m = 0) \wedge ((i \oplus k) \otimes \bar{m} = 0)] \end{aligned}$$

The condition in the square brackets can only be satisfied when $i = k \otimes \bar{m}$, and in this case $i \oplus k = (k \otimes \bar{m}) \oplus k = k \otimes m$. Thus,

$$\hat{y}_k = \sqrt{n} \sum_m \frac{\chi_m + \chi_{\bar{m}}}{2} \hat{x}_{k \otimes m} \hat{x}_{k \otimes \bar{m}}$$

■

Theorem 8.2 (*The crossover heuristic for schema in the Walsh basis.*) Let $x \in \Lambda$ and let \hat{y} denote $\mathcal{C}(x)$ expressed in the Walsh basis. Then

$$\hat{y}_k^{(u)} = \sum_m \frac{\chi_m + \chi_{\bar{m}}}{2} \hat{x}_{k \otimes m}^{(u \otimes m)} \hat{x}_{k \otimes \bar{m}}^{(u \otimes \bar{m})}$$

Proof.

$$\begin{aligned}
\hat{y}_k^{(u)} &= 2^{\#\bar{u}/2} \hat{y}_k \\
&= 2^{\#\bar{u} + \#u/2} \sum_m \frac{\chi_m + \chi_{\bar{m}}}{2} \hat{x}_{k \otimes m} \hat{x}_{k \otimes \bar{m}} \\
&= 2^{\#\bar{u} + \#u/2} \sum_m \frac{\chi_m + \chi_{\bar{m}}}{2} \left(2^{-\#(\bar{u} \otimes m)/2} \hat{x}_{k \otimes m}^{(u \otimes m)} \right) \left(2^{-\#(\bar{u} \otimes \bar{m})/2} \hat{x}_{k \otimes \bar{m}}^{(u \otimes \bar{m})} \right)
\end{aligned}$$

Consider the exponents:

$$\begin{aligned}
-\#(\bar{u} \otimes m)/2 - \#(\bar{u} \otimes \bar{m})/2 &= -\ell/2 + \#(u \otimes m)/2 - \ell/2 + \#(u \otimes \bar{m})/2 \\
&= -\ell + \#u/2
\end{aligned}$$

Thus,

$$\begin{aligned}
\hat{y}_k^{(u)} &= 2^{\#\bar{u} + \#u/2} 2^{-\ell + \#u/2} \sum_m \frac{\chi_m + \chi_{\bar{m}}}{2} \hat{x}_{k \otimes m}^{(u \otimes m)} \hat{x}_{k \otimes \bar{m}}^{(u \otimes \bar{m})} \\
&= \sum_m \frac{\chi_m + \chi_{\bar{m}}}{2} \hat{x}_{k \otimes m}^{(u \otimes m)} \hat{x}_{k \otimes \bar{m}}^{(u \otimes \bar{m})}
\end{aligned}$$

■

To continue the numerical example, suppose that 1-point crossover with crossover rate $1/2$ is applied to the $\ell = 5$ population for which $\hat{s}^{(10)} = \frac{1}{20} < 10, -2, 5, 1 >$. We want to compute $\hat{y}_k^{(10)}$ for $k = 0, 2, 8, 10$. For $k = 0, 2, 8$, for every crossover mask m , either $k \otimes m = 0$ or $k \otimes \bar{m} = 0$, so $\hat{y}_k^{(10)} = \hat{s}_k^{(k)} \hat{s}_0^{(10 \oplus k)} = \hat{s}_k^{(10)}$.

For $k = 10$, there are four possible nontrivial crossover masks, each with probability $1/8$. For two of these, $k \otimes m \neq 0$ and $k \otimes \bar{m} \neq 0$. This gives

$$\hat{y}_{10}^{(10)} = \frac{3}{4} \hat{s}_{10}^{(10)} + \frac{1}{4} \hat{s}_2^{(2)} \hat{s}_8^{(2)} = \frac{3}{4} \hat{s}_{10}^{(10)} + \frac{1}{4} \left(\sqrt{2} \hat{s}_2^{(10)} \right) \left(\sqrt{2} \hat{s}_8^{(10)} \right) = \frac{3}{4} \cdot \frac{1}{20} + \frac{1}{4} \cdot 2 \cdot \frac{-2}{20} \cdot \frac{5}{20} = \frac{3}{80} - \frac{1}{80} = \frac{1}{40}$$

Thus, $\hat{y}^{(10)} = \frac{1}{40} < 20, -4, 10, 1 >$.

The following theorem gives a simple formula for the exact change in the expected schema averages after crossover. It is a restatement of theorem ?? of [SW97] and theorem ?? of [SWA97]. It can also be easily derived from theorem 19.2 of [Vos99b] by setting the mutation rate to be zero.

Theorem 8.3 (*Exact schema theorem for crossover.*) *Let x be a population, and let $y = \mathcal{C}(x)$. Then*

$$y_k^{(u)} = \sum_m \frac{\chi_m + \chi_{\bar{m}}}{2} x_{k \otimes m}^{(u \otimes m)} x_{k \otimes \bar{m}}^{(u \otimes \bar{m})} \quad (3)$$

Proof.

$$\begin{aligned}
y_k^{(u)} &= 2^{-\#u/2} \sum_{v \in \Omega_u} (-1)^{k^T v} \hat{y}_v^{(u)} \\
&= 2^{-\#u/2} \sum_{v \in \Omega_u} (-1)^{k^T v} \sum_m \frac{\chi_m + \chi_{\bar{m}}}{2} \hat{x}_{v \otimes m}^{(u \otimes m)} \hat{x}_{v \otimes \bar{m}}^{(u \otimes \bar{m})} \\
&= 2^{-\#u/2} \sum_{i \in \Omega_{u \otimes m}} \sum_{j \in \Omega_{u \otimes \bar{m}}} \sum_m \frac{\chi_m + \chi_{\bar{m}}}{2} (-1)^{k^T (i \oplus j)} \hat{x}_{(i \oplus j) \otimes m}^{(u \otimes m)} \hat{x}_{(i \oplus j) \otimes \bar{m}}^{(u \otimes \bar{m})} \\
&= 2^{-\#u/2} \sum_m \frac{\chi_m + \chi_{\bar{m}}}{2} \sum_{i \in \Omega_{u \otimes m}} (-1)^{k^T i} \hat{x}_i^{(u \otimes m)} \sum_{j \in \Omega_{u \otimes \bar{m}}} (-1)^{k^T j} \hat{x}_j^{(u \otimes \bar{m})} \\
&= 2^{-\#u/2} \sum_m \frac{\chi_m + \chi_{\bar{m}}}{2} 2^{\#(u \otimes m)/2} x_{k \otimes m}^{(u \otimes m)} 2^{\#(u \otimes \bar{m})/2} x_{k \otimes \bar{m}}^{(u \otimes \bar{m})} \\
&= \sum_m \frac{\chi_m + \chi_{\bar{m}}}{2} x_{k \otimes m}^{(u \otimes m)} x_{k \otimes \bar{m}}^{(u \otimes \bar{m})}
\end{aligned}$$

■

Theorems 8.1, 8.2, and 8.3 show that the effect of crossover using mask m is to move the population towards linkage equilibrium relative to m . Following the population biologists (see [CK70] for example), we define the population x to be in *linkage equilibrium* relative to mask m if $\hat{x}_k = \hat{x}_{k \otimes m} \hat{x}_{k \otimes \bar{m}}$, or equivalently if $x_k^{(u)} = x_{k \otimes m}^{(u \otimes m)} x_{k \otimes \bar{m}}^{(u \otimes \bar{m})}$ for all $k \in \Omega$. If a population is in linkage equilibrium with respect to all masks of a family of crossover masks that separates any pair of bit positions, then the population will be completely determined by the order 1 schemata averages (or equivalently the Walsh coefficients \hat{x}_k with $\#k = 1$). This is formalized in theorem 3.0 of [VW98b] (Geiringer's theorem).

Continuing the numerical example, suppose that one-point crossover with crossover rate $1/2$ is applied to the $\ell = 5$ population whose schema averages for $u = 10$ are given by $s^{(10)} = \frac{1}{20} < 7, 8, 1, 4 >$. To apply theorem 8.3, we need $s^{(2)}$ and $s^{(8)}$. These are easily obtained from $s^{(10)}$: $s_0^{(2)} = s_0^{(10)} + s_8^{(10)} = \frac{2}{5}$, $s_2^{(2)} = s_2^{(10)} + s_{10}^{(10)} = \frac{3}{5}$, $s_0^{(8)} = s_0^{(10)} + s_2^{(10)} = \frac{3}{4}$, $s_8^{(8)} = s_8^{(10)} + s_{10}^{(10)} = \frac{1}{4}$.

Let $y = \mathcal{C}(s)$. As before, the probability of a crossover mask for which $u \otimes m \neq 0$ and $u \otimes \bar{m} \neq 0$ is $1/4$. Thus,

$$\begin{aligned}
y_0^{(10)} &= \frac{3}{4} s_0^{(10)} + \frac{1}{4} s_0^{(2)} s_0^{(8)} = \frac{3}{4} \cdot \frac{7}{20} + \frac{1}{4} \cdot \frac{2}{5} \cdot \frac{3}{4} = \frac{27}{80} \\
y_2^{(10)} &= \frac{3}{4} s_2^{(10)} + \frac{1}{4} s_2^{(2)} s_0^{(8)} = \frac{3}{4} \cdot \frac{8}{20} + \frac{1}{4} \cdot \frac{3}{5} \cdot \frac{3}{4} = \frac{33}{80} \\
y_8^{(10)} &= \frac{3}{4} s_8^{(10)} + \frac{1}{4} s_0^{(2)} s_8^{(8)} = \frac{3}{4} \cdot \frac{1}{20} + \frac{1}{4} \cdot \frac{2}{5} \cdot \frac{1}{4} = \frac{5}{80} = \frac{1}{16} \\
y_{10}^{(10)} &= \frac{3}{4} s_{10}^{(10)} + \frac{1}{4} s_2^{(2)} s_8^{(8)} = \frac{3}{4} \cdot \frac{4}{20} + \frac{1}{4} \cdot \frac{3}{5} \cdot \frac{1}{4} = \frac{15}{80} = \frac{3}{16}
\end{aligned}$$

One can check that $y^{(10)}$ is the Walsh transform of $\hat{y}^{(10)}$ computed earlier.

Corollary 8.4 (*Approximate schema theorem for crossover.*) *Let x be a population, and let $y =$*

$\mathcal{C}(x)$. Then

$$y_k^{(u)} \geq x_k^{(u)} \sum_m \frac{\chi_m + \chi_{\overline{m}}}{2} [(u \otimes m = u) \vee (u \otimes \overline{m} = u)]$$

Note that the summation over m includes just those crossover masks that do not “split” the mask u .

Proof. For u such that $u \otimes m = u$, we have:

$$\begin{aligned} u \otimes \overline{m} &= 0 & \text{and} \\ x_{k \otimes \overline{m}}^{(u \otimes \overline{m})} &= x_0^{(0)} = 1 & \text{and} \\ x_{k \otimes m}^{(u \otimes m)} &= x_k^{(u)} \end{aligned}$$

Similarly, for u such that $u \otimes \overline{m} = u$, we have:

$$\begin{aligned} u \otimes m &= 0 & \text{and} \\ x_{k \otimes m}^{(u \otimes m)} &= x_0^{(0)} = 1 & \text{and} \\ x_{k \otimes \overline{m}}^{(u \otimes \overline{m})} &= x_k^{(u)} \end{aligned}$$

Those terms in the summation of equation (3) for which $(u \oplus m = u) \vee (u \oplus \overline{m} = u)$ is not true are nonnegative. Thus, if we drop those terms from the summation, we get the equation of the corollary. \blacksquare

Corollary 8.5 (*Holland’s approximate schema theorem for 1-point crossover.*) *Let x be a population, and let $y = \mathcal{C}(x)$, where \mathcal{C} is defined through 1-point crossover with a crossover rate of c . Then*

$$y_k^{(u)} \geq x_k^{(u)} \left(1 - c \frac{\mathcal{L}(u)}{\ell - 1} \right)$$

Proof. One-point crossover can be defined using ℓ crossover masks with a nonzero probability. The crossover mask 0 has probability $1 - c$, and the masks of the form $2^i - 1$, $i = 1, \dots, \ell - 1$ have probability $c/(\ell - 1)$. The number of crossover masks such that $(u \otimes m = u) \vee (u \otimes \overline{m} = u)$ is not true is $\mathcal{L}(u)$. Thus, the probability that $(u \otimes m = u) \vee (u \otimes \overline{m} = u)$ is true is

$$\left(1 - c \frac{\mathcal{L}(u)}{\ell - 1} \right)$$

\blacksquare

It is not hard to give similar approximate schema theorems for other forms of crossover, such as two-point crossover and uniform crossover.

9 Mutation

In the Vose model, mutation is defined by means of mutation masks. If $j \in \Omega$, then the result of mutating j using a mutation mask $m \in \Omega$ is $j \oplus m$. The mutation heuristic is defined by giving

a probability distribution $\mu \in \Lambda$ over mutation masks. In other words, μ_m is the probability that $m \in \Omega$ is used. Given a population $x \in \Lambda$, the mutation heuristic $\mathcal{U} : \Lambda \rightarrow \Lambda$ is defined by

$$\mathcal{U}_k(x) = \sum_{j \in \Omega} \mu_{j \oplus k} x_j$$

The mutation heuristic is a linear operator: it can be defined as multiplication by the matrix U , where $U_{j,k} = \mu_{j \oplus k}$. In other words, $\mathcal{U}(x) = Ux$.

In the Walsh basis, the mutation heuristic is represented by a diagonal matrix.

Lemma 9.1 *The k th component of the mutation heuristic in the Walsh basis is given by*

$$\sqrt{n} \hat{\mu}_k \hat{x}_k$$

where $n = 2^\ell$.

Proof. It is sufficient to show that the Walsh transform \widehat{U} of U is diagonal since

$$\widehat{\mathcal{U}(x)} = WUx = (WUW)(Wx) = \widehat{U}\widehat{x}$$

The following shows that \widehat{U} is diagonal.

$$\begin{aligned} \widehat{U}_{j,k} &= \frac{1}{n} \sum_{v,w} (-1)^{j^T v} (-1)^{k^T w} U_{v,w} \\ &= \frac{1}{n} \sum_v \sum_w (-1)^{j^T v + k^T w} \mu_{v \oplus w} \end{aligned}$$

We now do a change of variable. Let $u = v \oplus w$, which implies that $w = v \oplus u$.

$$\begin{aligned} \widehat{U}_{j,k} &= \frac{1}{n} \sum_v \sum_u (-1)^{j^T v + k^T (v \oplus u)} \mu_u \\ &= \frac{1}{n} \sum_v \sum_u (-1)^{(j \oplus k)^T v + k^T u} \mu_u \\ &= \frac{1}{n} \sum_u (-1)^{k^T u} \mu_u \sum_v (-1)^{(j \oplus k)^T v} \\ &= \sum_u (-1)^{k^T u} \mu_u \\ &= \sqrt{n} \hat{\mu}_k [j = k] \end{aligned}$$

■

Define $\mu_k^{(u)} = \sum_{j \in \Omega_u} \mu_{k \oplus j}$. Theorem 7.1 shows that $\mu^{(u)} = W^{(u)} \hat{\mu}^{(u)}$, where $\hat{\mu}_k = 2^{\#u/2} \hat{\mu}_k$ for all $k \in \Omega_u$.

Define the $2^{\#u} \times 2^{\#u}$ matrix $U^{(u)}$ by $U_{j,k}^{(u)} = \mu_{j \oplus k}^{(u)}$. Note that $U = U^{(1)}$. The proof of lemma 9.1 shows that the Walsh transform $\widehat{U^{(u)}}$ of $U^{(u)}$ is diagonal and $\widehat{U^{(u)}}_{k,k} = 2^{\#u/2} \hat{\mu}_k$. Thus, it is consistent to write $\widehat{U}^{(u)}$ for $\widehat{U^{(u)}}$.

We now assume that each string position i , $i = 0, 1, \dots, \ell - 1$, is mutated independently of other positions: with a probability of p_i , the bit at position i is flipped. If all of the p_i are equal to a common value p , then p is called the *mutation rate*.

Under this assumption, the probability distribution for mutation masks is given by

$$\mu_m = \prod_{i=0}^{\ell-1} p_i^{m_i} (1 - p_i)^{1-m_i} \quad (4)$$

where m_i denotes bit i of m , and where 0^0 is interpreted to be 1. For example, the distribtuion for $\ell = 2$ is the vector

$$< (1 - p_0)(1 - p_1) \quad p_0(1 - p_1) \quad (1 - p_0)p_1 \quad p_0p_1 >^T$$

We now want to show that there is an equation similar to (4) for $\mu_m^{(u)}$. The next lemma is a step in that direction. For $u \in \Omega$, define $I(u) = \{i : 0 \leq i < \ell \text{ and } u_i = 1\}$.

Lemma 9.2 *For $v \in \Omega$,*

$$\sum_{k \in \Omega_u} \prod_{i \in I(u)} p_i^{k_i} (1 - p_i)^{1-k_i} = 1 \quad (5)$$

Proof. The proof is by induction on $\#u$.

If $\#u = 1$, then $u = 2^j$ for some j , and $I(u) = j$. Also, $\Omega_u = \{0, u\}$. Thus, the left side of equation (5) is $p_j^0(1 - p_j)^1 + p_j^1(1 - p_j)^0 = (1 - p_j) + p_j = 1$.

If $\#u > 1$, let $u = v \oplus w$ with $v \otimes w = 0$, $\#v > 0$, $\#w > 0$. Then

$$\begin{aligned} \sum_{k \in \Omega_u} \prod_{i \in I(u)} p_i^{k_i} (1 - p_i)^{1-k_i} &= \sum_{k \in \Omega_v} \sum_{r \in \Omega_w} \left(\prod_{i \in I(v)} p_i^{k_i} (1 - p_i)^{1-k_i} \right) \left(\prod_{j \in I(w)} p_j^{r_j} (1 - p_j)^{1-r_j} \right) \\ &= \left(\sum_{k \in \Omega_v} \prod_{i \in I(v)} p_i^{k_i} (1 - p_i)^{1-k_i} \right) \left(\sum_{r \in \Omega_w} \prod_{j \in I(w)} p_j^{r_j} (1 - p_j)^{1-r_j} \right) \\ &= 1 \end{aligned}$$

■

Lemma 9.3 *For $u \in \Omega$ and $m \in \Omega_u$,*

$$\mu_m^{(u)} = \prod_{i \in I(u)} p_i^{m_i} (1 - p_i)^{1-m_i} \quad (6)$$

Proof.

$$\begin{aligned}
\mu_m^{(u)} &= \sum_{v \in \Omega_{\bar{u}}} \mu_{m \oplus v} \\
&= \sum_{v \in \Omega_{\bar{u}}} \left(\prod_{i \in I(u)} p_i^{(m \oplus v)_i} (1 - p_i)^{1 - (m \oplus v)_i} \right) \left(\prod_{j \in I(\bar{u})} p_j^{(m \oplus v)_j} (1 - p_j)^{1 - (m \oplus v)_j} \right) \\
&= \left(\prod_{i \in I(u)} p_i^{m_i} (1 - p_i)^{1 - m_i} \right) \left(\sum_{v \in \Omega_{\bar{u}}} \prod_{j \in I(\bar{u})} p_j^{(m \oplus v)_j} (1 - p_j)^{1 - (m \oplus v)_j} \right)
\end{aligned}$$

Do a change of variable: let $w = v \oplus (m \otimes \bar{u})$. Then

$$\begin{aligned}
\sum_{v \in \Omega_{\bar{u}}} \prod_{j \in I(\bar{u})} p_j^{(m \oplus v)_j} (1 - p_j)^{1 - (m \oplus v)_j} &= \sum_{w \in \Omega_{\bar{u}}} \prod_{j \in I(\bar{u})} p_j^{w_j} (1 - p_j)^{1 - w_j} \\
&= 1
\end{aligned}$$

■

The next step is to compute the Walsh transform of the mutation probability distribution under this assumption. It is helpful to do a change of coordinates. For each $i = 0, 1, \dots, \ell - 1$, let $q_i = 1 - 2p_i$. Under this change of coordinates, equation (6) is equivalent to

$$\mu_m^{(u)} = 2^{-\#u} \prod_{i \in I(u)} (1 + (1 - 2m_i)q_i)$$

Lemma 9.4 For $m \in \Omega_u$,

$$\hat{\mu}_m^{(u)} = 2^{-\#u/2} \prod_{i \in I(m)} q_i$$

Proof. The proof is by induction on $\#u$. For the base case, assume that $\#u = 1$. Then $u = 2^i$ for some i , and

$$\begin{aligned}
\hat{\mu}^{(u)} &= W^{(u)} \mu^{(u)} \\
&= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \frac{1}{2} \begin{bmatrix} 1 + q_i \\ 1 - q_i \end{bmatrix} \\
&= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ q_i \end{bmatrix}
\end{aligned}$$

For $\#u > 1$, we have

$$\hat{\mu}_m^{(u)} = 2^{-\#u/2} \sum_{j \in \Omega_u} (-1)^{m^T j} 2^{-\#u} \prod_{i \in I(u)} (1 + (1 - 2j_i)q_i)$$

Let $u = v \oplus w$ where $v \otimes w = 0$, $v \neq 0$, and $w \neq 0$.

$$\begin{aligned}
\hat{\mu}_m^{(u)} &= 2^{-\frac{3}{2}\#(v \oplus w)} \sum_{j \in \Omega_{v \oplus w}} (-1)^{m^T j} \prod_{i \in I(v \oplus w)} (1 + (1 - 2j_i)q_i) \\
&= \left(2^{-\frac{3}{2}\#v} \sum_{j \in \Omega_v} (-1)^{(m \otimes v)^T j} \prod_{i \in I(v)} (1 + (1 - 2j_i)q_i) \right) \\
&\quad \left(2^{-\frac{3}{2}\#w} \sum_{j \in \Omega_w} (-1)^{(m \otimes w)^T j} \prod_{i \in I(w)} (1 + (1 - 2j_i)q_i) \right) \\
&= \left(\hat{\mu}_{m \otimes v}^{(v)} \right) \left(\hat{\mu}_{m \otimes w}^{(w)} \right) \\
&= \left(2^{-\#v/2} \prod_{i \in I(m \otimes v)} q_i \right) \left(2^{-\#w/2} \prod_{i \in I(m \otimes w)} q_i \right) \\
&= 2^{-\#u/2} \prod_{i \in I(m)} q_i
\end{aligned}$$

■

Lemma 9.5

$$\hat{\mu}_m = \frac{1}{\sqrt{n}} \prod_{i \in I(m)} q_i$$

Proof.

$$\hat{\mu}_m = \hat{\mu}_m^{(1)} = 2^{\#1/2} \prod_{i \in I(m)} q_i = \frac{1}{\sqrt{n}} \prod_{i \in I(m)} q_i$$

■

Theorem 9.6 (*The mutation heuristic in the Walsh Basis.*) Let $x \in \Lambda$ be a population, and let $y = \mathcal{U}(x)$. If $k \in \Omega_u$, then

$$\hat{y}_k^{(u)} = \hat{x}_k^{(u)} \prod_{i \in I(k)} q_i$$

Proof.

$$\begin{aligned}
\hat{y}_k^{(u)} &= 2^{\#\bar{u}/2} \hat{y}_k \\
&= 2^{\#\bar{u}/2} 2^{\ell/2} \hat{\mu}_k \hat{x}_k \quad \text{by lemma 9.1} \\
&= 2^{\#\bar{u}/2} 2^{\ell/2} \left(2^{-\#\bar{u}/2} \mu_k^{(u)} \right) \left(2^{-\#\bar{u}/2} x_k^{(u)} \right) \\
&= 2^{\#u/2} \left(2^{-\#u/2} \prod_{i \in I(k)} q_i \right) \hat{x}_k^{(u)} \\
&= \hat{x}_k^{(u)} \prod_{i \in I(k)} q_i
\end{aligned}$$

■

Theorem 9.6 shows how mutation affects a population. If $k \neq 0$, and if for every i , $0 < p_i \leq 1/2$, then $\prod_{i \in I(k)} q_i < 1$. Thus, $|\hat{y}_k^{(u)}| < |\hat{x}_k^{(u)}|$. Mutation is decreasing the magnitude of the schema Walsh coefficients (except for the index 0 coefficient which is constant at $2^{-\#u/2}$). If all of these Walsh coefficients were zero, then Theorem 7.1 shows that all of the corresponding schema averages would be equal. In other words, mutation drives the population towards uniformity.

To continue the numerical example, we take apply mutation with a mutation rate of $1/8$ to the population y of the previous section. We start with $\hat{y}^{(10)} = \frac{1}{40} < 20, -4, 10, 1 >$. Let $z = \mathcal{U}(y)$. For all i , $q = q_i = 1 - 2p_i = 1 - 1/4 = 3/4$. Thus,

$$\begin{aligned}\hat{z}_0^{(10)} &= \hat{y}_0^{(10)} = \frac{1}{2} \\ \hat{z}_2^{(10)} &= \hat{y}_2^{(10)} \cdot q = \frac{-1}{10} \cdot \frac{3}{4} = -\frac{3}{40} \\ \hat{z}_8^{(10)} &= \hat{y}_8^{(10)} \cdot q = \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{16} \\ \hat{z}_{10}^{(10)} &= \hat{y}_{10}^{(10)} \cdot q^2 = \frac{1}{40} \cdot \frac{9}{16} = \frac{9}{640}\end{aligned}$$

Lemma 9.7 For $u \in \Omega$,

$$\hat{y}^{(u)} = \hat{U}^{(u)} \hat{x}^{(u)}$$

Proof. This is just a rewriting of the equation of theorem 9.6 into matrix form. ■

The following theorem can be easily derived from theorem 19.2 of [Vos99b] by setting the crossover rate to be zero.

Theorem 9.8 (*The exact schema theorem for mutation.*) Let $x \in \Lambda$ be a population, and let $y = \mathcal{U}(x)$ where \mathcal{U} corresponds to mutating bit i with probability p_i for $i = 0, 1, \dots, \ell - 1$. Then

$$y^{(u)} = U^{(u)} x^{(u)} \tag{7}$$

Proof.

$$\begin{aligned}y^{(u)} &= W^{(u)} \hat{y}^{(u)} \\ &= W^{(u)} \hat{U}^{(u)} \hat{x}^{(u)} \\ &= W^{(u)} (W^{(u)} U^{(u)} W^{(u)}) W^{(u)} x^{(u)} \\ &= U^{(u)} x^{(u)}\end{aligned}$$

■

We continue the numerical example. We start with the schema averages computed in the crossover section: $y^{(10)} = \frac{1}{80} < 27, 33, 5, 15 >$ and let $z = \mathcal{U}(y)$ where \mathcal{U} corresponds to mutation with a mutation rate of $1/8$. Recall that $\mu^{(u)}$ is given by equation (6), so

$$\mu^{(10)} = < (1-p)^2, p(1-p), (1-p)p, p^2 > = \frac{1}{64} < 49, 7, 7, 1 >$$

The entries of $U^{(u)}$ are given by $U_{j,k}^{(u)} = \mu_{j \oplus k}^{(u)}$, so

$$z^{(10)} = U^{(10)} y^{(10)} = \frac{1}{64} \begin{bmatrix} 49 & 7 & 7 & 1 \\ 7 & 49 & 1 & 7 \\ 7 & 1 & 49 & 7 \\ 1 & 7 & 7 & 49 \end{bmatrix} \cdot \frac{1}{80} \begin{bmatrix} 27 \\ 33 \\ 5 \\ 15 \end{bmatrix} = \frac{1}{1280} \begin{bmatrix} 401 \\ 479 \\ 143 \\ 257 \end{bmatrix}$$

Corollary 9.9 (*The approximate schema theorem for mutation.*) *Let $x \in \Lambda$ be a normalized population, and let $y = \mathcal{U}(x)$. Assume that \mathcal{U} corresponds to mutation where each bit is mutated (flipped) with probability p . Then*

$$y_k^{(u)} \geq (1-p)^{\#u} x_k^{(u)}$$

Proof. The diagonal entries of $U^{(u)}$ are all equal to $\hat{\mu}_0^{(u)} = \prod_{i \in I(u)} (1-p_i)$. Under the assumption of this corollary, $p_i = p$ for all i , so the diagonal entries of $U^{(u)}$ are all equal to $(1-p)^{\#u}$. The off-diagonal entries of $U^{(u)}$ are all nonnegative. If we drop the off-diagonal entries in the computation of equation (7), we get the result of this corollary. ■

10 Computational Complexity

In this section we give the computational complexity of computing the schema averages for a family of competing schema averages after one generation of the simple GA.

It is more efficient to compute the schema averages after selection using the normal basis using the algorithm given in section 5, convert to the Walsh basis using the Fast Walsh transform (see Appendix A), compute the effects of crossover and mutation in the Walsh basis, and convert back to normal coordinates using the fast Walsh transform. To convert from $x^{(u)}$ to $\hat{x}^{(u)}$ by the fast Walsh transform has complexity $\Theta(\#u \cdot 2^{\#u})$ ([Vos99b]). The complexity of the computation of theorem 8.2 is $\Theta(\#u \cdot 2^{\#u})$ for one or two point crossover (since the summation over m is $\Theta(\#u)$). The complexity of the computation of theorem 9.6 is also $\Theta(\#u \cdot 2^{\#u})$. Thus, the overall computational complexity (assuming an initial finite population and one or two point crossover) is $\Theta(\#u \cdot 2^{\#u} + rK)$ where K was defined as the cost of doing one function evaluation. Note that the only dependence on the string length is through K . Thus, it is possible to compute schema averages exactly for very long string lengths.

11 Conclusion

We have given a version of the Vose infinite population model where the crossover heuristic function and the mutation heuristic function are separate functions, rather than combined into a single mixing heuristic function.

We have shown how the expected behavior of a simple genetic algorithm relative to a family of competing schemata can be computed exactly over one generation.

As was mentioned in section 7, these schema averages over a family of competing schemata correspond to a coordinate subspace of Λ as expressed in the Walsh basis. In [VW98a], it was shown that the mixing (crossover and mutation) heuristic is invariant over coordinate subspaces in the Walsh basis. We have explicitly shown how the Vose infinite population model (the heuristic function \mathcal{G}) can be computed on these subspaces. In fact, the model works in essentially the same way on schema averages as it does on individual strings.

The formulas are simply stated and easy to understand. They are computationally feasible to apply even for very long string lengths if the order of the family of competing schemata is small. (The formulas are exponential in the order of the schemata.)

A result like the exact schema theorem is most useful if it can be applied over multiple generations. The results of this paper show that the obstacle to doing this is selection, rather than crossover and mutation. The result of the exact schema theorem is the exact schema averages of the family of competing schemata (or the corresponding Walsh coefficients) after one generation. These correspond to an “infinite population” which has nonzero components over all elements of Ω . If the string length is long and no assumptions are made about the fitness function, then the effect of selection on the schema averages for the next generation will be computationally infeasible to compute. Thus, in order to apply the exact schema theorem over multiple generations for practically realistic string lengths, one will have to make assumptions about the fitness function. A subsequent paper will explore this problem.

Acknowledgements: The author would like to thank Yong Zhao, who proofread a version of this paper.

References

- [Alt95] Lee Altenberg. The schema theorem and Price’s theorem. In L. Darrell Whitley and Michael D. Vose, editors, *Foundations of genetic algorithms 3*, pages 23–49. Morgan Kaufmann, 1995.
- [BG87] C. L. Bridges and D. E. Goldberg. An analysis of reproduction and crossover in a binary-coded genetic algorithm. In J. Grefenstette, editor, *Proceedings of the Second International Conference on Genetic Algorithms*, pages 9–13, Hillsdale, N. J., 1987. Lawrence Erlbaum Associates.

- [CK70] James F. Crow and Motoo Kimura. *An Introduction to Population Genetics Theory*. Burgess Publishing Company, Minneapolis, Minnesota, 1970.
- [Hol75] John Holland. *Adapdation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, Michigan, 1975.
- [SW97] C. R. Stephens and H. Waelbroeck. Effective degrees of freedom in genetic algorithms and the block hypothesis. In Thomas Bäck, editor, *Proceedings of the Seventh International Conference on Genetic Algorithms*, pages 34–40, San Mateo, 1997. Morgan Kaufman.
- [SWA97] C. R. Stephens, H. Waelbroeck, and R. Aguirre. Schemata as building blocks: does size matter. In *Foundations of Genetic Algorithms (FOGA-5)*, pages 117–133, San Mateo, 1997. Morgan Kaufmann.
- [VL91] M. D. Vose and G. E. Liepins. Punctuated equilibria in genetic search. *Complex Systems*, 5:31–44, 1991.
- [Vos96] M. D. Vose. Modeling simple genetic algorithms. *Evolutionary Computation*, 3(4):453–472, 1996.
- [Vos99a] M. D. Vose. Random heuristic search. *submitted to Theoretical Computer Science*, 1999.
- [Vos99b] M. D. Vose. *The Simple Genetic Algorithm: Foundations and Theory*. MIT Press, Cambridge, MA, 1999.
- [VW98a] M. D. Vose and A. H. Wright. The simple genetic algorithm and the Walsh transform: Part I, theory. *Evolutionary Computation*, 6(3):253–273, 1998.
- [VW98b] M. D. Vose and A. H. Wright. The simple genetic algorithm and the Walsh transform: Part II, the inverse. *Evolutionary Computation*, 6(3):275–289, 1998.

Appendix: Table of Notation

$[e]$	$= 1$ if e is true, 0 if e is false
ℓ	The string length
c	The arity of the alphabet used in the string representation
Ω	The set of binary strings of length ℓ
n	$= c^\ell$, the number of elements of Ω
r	The population size
$u \oplus v$	The strings j and k are bitwise added mod 2, (or bitwise XORed)
$u \otimes v$	The strings j and k are bitwise multiplied mod 2, (or bitwise ANDed)
\bar{u}	The ones complement of the string j
$\#u$	The number of ones in the binary string u
$k^T j$	The same as $\#(k \otimes j)$, the number of ones in $k \otimes j$
Λ	The set of nonnegative real-valued vectors indexed over Ω whose sum is 1 $=$ the set of normalized populations $=$ the set of probability distributions over Ω
Ω_u	$= \{k \in \Omega : u \otimes k = k\}$
$\Omega_{\bar{u}} \oplus v$	$= \{j \oplus v : j \in \Omega_{\bar{u}}\}$ = the schema with fixed positions masked by u and specified by v
$x_v^{(u)}$	$= \sum_{j \in \Omega_{\bar{u}}} x_{j \oplus v}$ (assuming that $\sum_j x_j = 1$). The schema average or sum for the schema $\Omega_{\bar{u}} \oplus v$
$x^{(u)}$	The vector of schema averages for the family of schemata $\{\Omega_{\bar{u}} \oplus v : v \in \Omega_u\}$
W	The Walsh transform matrix, indexed over $\Omega \times \Omega$. $W_{i,j} = \frac{1}{\sqrt{n}}(-1)^{i^T j}$
$W^{(u)}$	The Walsh transform matrix, indexed over $\Omega_u \times \Omega_u$. $W_{i,j}^{(u)} = 2^{-\#u/2}(-1)^{i^T j}$
\hat{x}	$= Wx$, the Walsh transform of normalized population x
$\hat{x}^{(u)}$	$= 2^{\#\bar{u}/2} \hat{x}$, also the Walsh transform $W^{(u)} x^{(u)}$ of $x^{(u)}$ with respect to Ω_u
χ_m	The probability that $m \in \Omega$ is used as a crossover mask
μ_m	The probability that $m \in \Omega$ is used as a mutation mask
$\mu_m^{(u)}$	$\sum_{j \in \Omega_{\bar{u}}} \mu_{k \oplus j}$
p_i	The probability that bit i is flipped in the mutation step
q_i	$= 1 - 2p_i$
U	The matrix indexed over $\Omega \times \Omega$ and defined by $U_{j,k} = \mu_{j \oplus k}$
$U^{(u)}$	The matrix indexed over $\Omega_u \times \Omega_u$ and defined by $U_{j,k}^{(u)} = \mu_{j \oplus k}^{(u)}$